UIUC iSchool 数据科学课程群调查研究*

■ 杨瑞仙 万佳琦
郑州大学信息管理学院 郑州 450001

摘 要：[目的/意义] 调查数据科学课程群建设现状，聚焦数据科学人才培养方案，为我国高校信息学院数据科学教学实践提供参考和借鉴。[方法/过程] 基于 UIUC(美国伊利诺伊大学香槟分校)信息科学学院的数据科学课程实践，首先调研该院数据科学相关课程的名称、简介、学制学时、授课形式、授课教师及授课对象，然后从培养对象类型、授课形式、授课合作程度和课程内容 4 个方面对课程群进行系统分类和比较分析，最后对我国高校数据科学课程建设提出若干建议。[结果/结论] UIUC 数据科学课程群可分为六大类别，面向本硕博各阶段学生，采用线上线下相结合的混合式教学方式，通过教师合作开展授课，教学内容紧密跟随数据科学岗位市场需求。因此，我国高校在数据科学领域应强化培育连续性、丰富教学创新性、加强教师授课合作性、增强研究方向完备性。

关键词：数据科学 UIUC iSchool 课程建设 课程改革
分类号：G251
DOI：10.13266/j.issn.0252-3116.2020.16.013

1 引言

第四次工业革命促进了社会发展方式的转变，大数据基础研究、产品研发和业务应用等各类人才较为短缺[1]。Talkingdata 发布的专业数据人才教育行业生态报告显示，到 2025 年我国将面临 200 万数据科学人才缺口[2]。现阶段国内外数据科学岗位技能要求呈现“精于一而又博学”的特点，高端人才市场需求量大且竞争激烈。工信部《大数据产业发展规划(2016-2020)》明确指出大数据产业人才队伍建设亟需加强。为顺应时代发展潮流和满足国家建设需求，2016 年教育部首次增设“数据科学与大数据技术”本科专业，仅 2019 年新增设的数据科学与大数据技术专业点高达 196 个，大数据管理与应用专业点达 25 个[3]。因此，探究数据科学课程建设现状，优化数据科学人才培养方案，建立满足社会市场需求的数据科学教育体系势在必行。美国伊利诺伊大学香槟分校(University of Illinois at Urbana-Champaign，简称“UIUC”)信息科学学院(School of Information Sciences，简称“iSchool”)一直是图情档领域和信息科学领域的领航者，自 1996 年以来，该院图书情报专业排名稳居全美第一。因此，在“双一流”学科建设背景下，笔者希望通过对 UIUC iSchool 数据科学课程开展情况的调研，为我国高校数据科学专业的课程建设和人才培养提供参考。

2 相关研究

通过文献调研发现，国外学者主要集中于数据科学学科理论研究、数据科学教育教学实践研究和数据科学的应用研究。在数据科学学科理论研究方面，知名丹麦籍计算机科学家、图灵奖获得者 P. Naur 于 1974 年率先正式提出术语数据科学(Data Science)。他在其专著《计算机方法的简明调查》(*Concise Survey of Computer Methods*)的前言中阐述了数据科学的内涵，并在书中辨析了数据科学与数据学(Datalogy)的差异[4]。随后，D. Conway 在 2010 年提出了数据科学韦恩图(The Data Science Venn Diagram)，首次确立了数据科学的学科地位——处于机器学习、数学与统计学和某一专业领域实务知识的交叉地带[5]。在数据科学教育教学实践研究方面，国外学者较多侧重研究某一典型院校面向本科生数据科学课程的具体情况。如

* 本文系郑州大学研究生教改项目“大数据背景下的数据挖掘课程改革研究”(项目编号：YJSJY201945)和郑州大学 2019 年度“一院一对标”本科教育教学建设项目“面向国际一流 iSchool 的图情档本科教育创新提升研究”研究成果之一。
作者简介：杨瑞仙(ORCID：0000-0002-5982-6339)，副院长，副教授，博士后，E-mail：yrx@zzu.edu.cn；万佳琦(ORCID：0000-0003-4005-0881)，硕士研究生。
收稿日期：2020-02-02 修回日期：2020-04-13 本文起止页码：122-131 本文责任编辑：杜杏叶

P. Anderson[6]等描述了美国南卡罗来纳州查尔斯顿学院(College of Charleston,South Carolina, USA)的数据科学课程计划及实施经验。B. Baumer 和 S. College[7]介绍了美国著名七姊妹学院之一史密斯学院的数据科学教学模块组成,如数据可视化、数据操作(Data Manipulation)/数据整理(Data wrangling)、计算统计、机器学习(或统计学习)以及拓展议题,如空间分析、文本挖掘、数据探索、网络科学等,也为学生数据处理能力的培养提供可行性建议。R. Veaux 和 M. Agarwal[8]等详细介绍了帕克城市数学学院(PCMI)2016 年暑期数据科学本科课程的制定指南,该指南旨在为学院规划数据科学专业提供一些结构性参考。V. Song 和 Y. Zhu[9]提供了一个分层的数据科学教育框架(Data Science Education Framework),由数据科学三大支柱(人员、技术和数据)、计算思维、数据驱动范式和数据科学生命周期四个模块组成。基于该框架,他们在德雷塞尔大学开展了基于用户、基于工具和基于应用程序的数据科学课程。此外,国外学者还从 iSchools 与数据科学教育的关系出发阐述领域总体概况,如 V. Song 和 Y. Zhu[10]认为 iSchools 是数据科学教育的中心枢纽,图情领域的跨学科师资团队能够培育出大量具备多种技能和广阔视野的数据科学家。在数据科学应用研究方面,国外研究较多涉及政府治理、业务管理、医疗保健、生命科学、金融经济、数据新闻等领域。

近几年,国内学者开始陆续对数据科学项目展开调查和研究,内容集中在国内外数据科学教学实践的对比分析。在数据科学课程建设方面,朝乐门、杨灿军[11]等调查了全球数据科学课程建设现状,总结了国内外数据科学课程的共性特色、共识经验及问题挑战,在探讨解决对策时重点提出了 10 个关于数据科学课程设计与教学改革的核心问题;之后,朝乐门、邢春晓等[12]从特色课程角度深入调研了 8 所世界一流名校数据科学专业的培养方案并对其特色课程进行分类,为我国大数据教育建设中存在的不足与曲解提供建议;李莎莎、周竞文等[13]分别从本科教育和硕士研究生教育层次对比分析了国内外 14 所高校的数据科学及大数据相关专业,结合各院校的课程设置和培养方案,为建设大数据人才培养模式提供建议。

在 iSchools 数据科学教育项目研究方面,闫慧、张钰浩等[14]调研了 iSchools 联盟中 10 所院校数据科学教育相关专业的 141 门课程信息,并将这些课程按照课程内容分为 12 类:基础理论课、相关学科理论基础课、统计学、机器学习、数据可视化、数据分析、数据科学工具、数据挖掘、数据库和数据管理、数据的社会影响、数据政策与法规和自主学习。苏日娜等[15]对开设数据科学研究生项目的 15 所 iSchools 高校从专业学科优势、学科体系划分、课程目标、核心课程设置、课程制度等方面进行研究,探讨了图书馆与信息科学(LIS)视域下数据科学学科建设及人才培养等问题。邓胜利、付少雄[16]从教育体系、学术研究与社会实践三个维度对德雷塞尔大学计算与信息学院图书情报学科建设情况进行调研,探究数据驱动以及计算科学与信息科学融合下的图情学科的新发展。

综上,国内 iSchools 数据科学教育相关研究仍处于起步阶段,研究方法较多采用网站调研,研究内容以对数据科学教学实践的宏观对比分析为主,体现在数据科学教育项目和课程建设方面;缺乏对国外一流代表性院校数据科学课程信息微观细节的深入挖掘和分析,缺乏对数据科学课程的实地考察研究,缺乏对数据科学课程教学细节的分维度调研。UIUC iSchool 作为国际图情学科的引领者,是郑州大学信息管理学院本科教育教学建设“一院一对标”项目的对标单位。因此本文将结合我院自身实际和国内教育实践,通过实地考察辅以网站调研,充分探究对标学院数据科学课程群开设情况,获取支撑数据科学课程建设和发展的一手资料,从培养对象类型、授课形式、授课合作程度和课程内容四个维度展开课程建设的微观剖析,以期对我院,乃至国内数据科学课程的建设提供参考和借鉴。

3　数据来源与研究方法

本研究的数据来源于实地考察和网站调研。2018 年 7 月至 2019 年 7 月本文作者之一杨瑞仙在 UIUC 访学期间共研修了四门数据科学课程——Data, Statistical Models, and Information(数据、统计模型和信息)、Information Organization & Access(信息组织与获取)、Foundations of Data Science(数据科学导论)、Theory and Practice Data Cleaning(数据清洗理论与实践),她通过线上线下相结合的形式参与所有课程的学习和讨论。2019 年 10 月本文两位作者共同在 UIUC iSchool 学院网站上(https://ischool. illinois. edu/),以“data”为关键词在网页搜索框中进行检索,选择“ Type = Course”后得到 42 条课程数据结果,并以此作为研究对象进行资料的梳理和分析。

在调研过程中,笔者紧密结合课堂实践收获、课程共享资料和网站课程信息,利用统计分析、对比分析和

归纳总结等多种研究方法，对 UIUC iSchool 的 42 门数据科学相关课程细节进行系统梳理和深入分析，重点考察研究对象的课程名称及简介、学制学时、授课形式、授课对象及师资情况。

4 UIUC 数据科学课程群总体概况

不同学者对数据科学有不同的定义，J. Stanton[17] 认为数据科学是与收集、准备、分析、可视化、管理和保存大批量数据相关工作的新兴领域；V. Dhar[18] 认为数据科学是从数据中获取知识的研究；F. Provost 和 T. Fawcett[19] 认为数据科学是通过自动分析数据来理解现存现象的原理、过程和技术。密歇根大学数据科学项目（Data Science Initiative，DSI）[20] 认为数据科学是将科学发现与实践相联结的一系列过程，它涉及大规模异构数据的收集、管理、处理、分析、可视化和解读，这些数据往往与可转化的、跨学科的科学应用相关。

通过归纳分析各种定义，不难发现数据科学的研究和应用对象是大批量数据，其基本流程包含数据的收集、整理、加工、展现等。2015 年，斯坦福大学统计学教授 D. L. Donoho[20] 在其数据科学 50 年的报告中明确指出，完整的数据科学可以划分为六大部分，分别是数据探索和准备、数据表示和转换、数据计算、数据建模、数据可视化和演示、数据科学相关科学，各部分表示如图 1 所示：

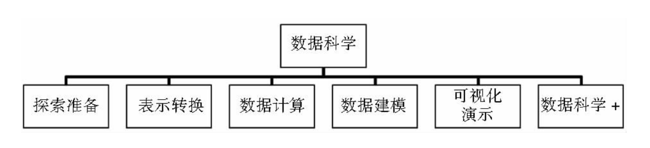


图 1 数据科学的组成[20]

笔者据此对 UIUC iSchool 数据科学课程进行了分类汇总和分析。表 1 显示了具体情况。其中，数据探索和准备类约占 16.67%，其核心课程有数据科学基础、数据管护基础、信息处理基础，重视数据库管理与体系结构方面的理论与实践，强化系统思想。数据表示和转换类约占 16.67%，核心课程有数据清洗理论与实践、元数据理论与实践，重视学生利用计算机处理数据的编程能力。数据可视化和演示类约占 9.52%，其核心课程包含数据可视化、数据科学故事化，关注沟通交流数据潜藏的信息、知识。数据计算类约占 4.76%，以云计算导论、机器学习社会计算前沿为重点内容。数据建模类约占 7.14%，核心课程是 Data，Statistical Models，and Information（数据、统计模型与信息）。“数据科学 +”学科类比例最高约占 45.24%，多为选修课，如竞争情报与知识管理、科研数据政策研讨、实用健康数据分析、数据伦理、生物信息学问题与研究、社交媒体分析、金融预测分析等，说明该学院课程涉及领域广泛、教学内容新颖，学生研修的可选择性强，有利于深化培养各类专项人才，使学生通过个性化定制课程获得应对就业环境压力的能力。

表 1 UIUC iSchool 数据科学课程名称及分类

课程类别	课程名称
数据探索和准备类（7 门，占 16.67%）	Data Mining（数据挖掘）；Foundations of Data Science（数据科学基础）；Advanced Data Science（高级数据科学）；Foundations of Data Curation（数据管护基础）；Big Data Infrastructures for Research and Development（用于研发的大数据基础设施）；Foundations of Information Processing（信息处理基础）；Database Administration and Scaling for IS（数据库管理及系统设计）
数据表示和转换类（7 门，占 16.67%）	Open Data Mashups（开放数据融合）；Theory and Practice Data Cleaning（数据清洗理论与实践）；Programming for Analytics and Data Processing（数据分析处理编程）；Introduction to Databases（数据库导论）；Database Design and Prototyping（数据库设计与原型构建）；Qualitative Methods Research（定性方法研究）；Metadata in Theory and Practice（元数据理论与实践）
数据计算类（2 门，占4.76%）	Introduction to Cloud Computing（云计算导论）；Advanced Topics in Machine Learning & Social Computing（机器学习和社会计算前沿）
数据可视化和演示类（4 门，占 9.52%）	Data Visualization（数据可视化）；Advanced Data Visualization（高级数据可视化）；Data Science Storytelling（数据科学故事化）；Network Analysis（网络分析）
数据建模类（3 门，占7.14%）	Methods for Data Science（数据科学方法概论）；Introduction to Data Science（数据科学导论）；Data，Statistical Models，and Information（数据、统计模型和信息）
数据科学 + 学科类（19 门，占 45.24%）	Community Data（社区数据）；Data Ethics（数据伦理）；Scientific Data Policy Seminar（科研数据政策研讨）；Practical Health Data Analytics（实用健康数据分析）；Data Warehousing and Business Intelligence（数据仓库与商务智能）；Data Science in the Humanities（人文中的数据科学）；Business Analytics（商务分析）；Internet of Things and Applications for Business（物联网及其商业应用）；Social Media Analytics（社交媒体分析）；Privacy in the Internet Age（互联网时代的隐私）；Information Ethics（信息伦理）；Local，Regional，and Global Intersections in LIS（图情领域内本地、区域、全球交叉分析）；Competitive Intelligence and Knowledge Management（竞争情报与知识管理）；Information Consulting（信息咨询）；Bioinformatics Problems and Research（生物信息学问题与研究）；Predictive Analysis in Finance（金融中的预测分析）；Copyright for Information Professionals（信息科研人员版权研究）；Digital Humanities（数字人文）；Professional Communication for Library and Other Information Professionals（图书馆及其信息研究人员的专业交流）

UIUC iSchool 教学形式个性多样，每年有三个学期，分别为春季学期(Spring)、秋季学期(Fall)和夏季小学期(Summer)。其中，春秋两个学期的课程设置几乎没有区别，仅个别课程授课教师有些调整，授课形式有变。所调研的数据科学课程群中在夏季小学期的授课课程分别是商务分析(Business Analytics)、信息处理基础(Foundations of Information Processing)、数据库基础(Introduction to Databases)、元数据理论与实践(Metadata in Theory and Practice)和信息研究人员的版权研究(Copyright for Information Professionals)。这五门夏季小学期课程中，商务分析(Business Analytics)、信息处理基础(Foundations of Information Processing)和元数据理论与实践(Metadata in Theory and Practice)同时也在春季学期和秋季学期开设；但剩余的两门数据库基础(Introduction to Databases)和信息研究人员的版权研究(Copyright for Information Professionals)仅在春季学期开设，秋季取消。这说明部分课程授课质量高和学生满意度高收获大，受到追捧。夏季小学期每门课程的总课时量约为 24－48 学时，相较于春秋季学期每门课程的总课时量 32－64 学时较短。在较短时间内，高效完成重点课程和热门课程的教学活动，有利于满足学生的学习需求，显著增加这些课程的影响。此外，该校的国际短期交流项目——全球教育与培训(Global Education and Training，GET)也是夏季小学期的特色，例如，2018 年 7 月来自中国一流高校的本科生通过 GET 的资助来到 UIUC 信息学院学习了由 J. Diesner 副教授讲授的网络分析课程(数据科学课程群中数据可视化和演示类课程之一)。

5 UIUC 数据科学课程群设置分析

在总体梳理 UIUC iSchool 数据科学课程的分类、数量分布、特色学期等内容基础上，还需要进一步从不同视角探究 UIUC 数据科学教育与教学活动的经验方法。具体来说，笔者主要从培养对象层次、授课形式、教师合作授课情况、授课内容 4 个维度展开分析。

5.1 培养对象层次分析

通过统计，笔者发现 UIUC iSchool 数据科学课程群中 91% 的授课对象是硕士生，7% 是本科生，2% 是博士生，现阶段硕士培养体系较为成熟(见图 2)。

本科生阶段由于学生数学统计学知识不够夯实，专业基础积淀不够深厚，很多数据科学课程无法按照数据科学流程及进度设置教学计划和课程计划；因此，UIUC iSchool 数据科学课程群中本科阶段仅涉及 3 门

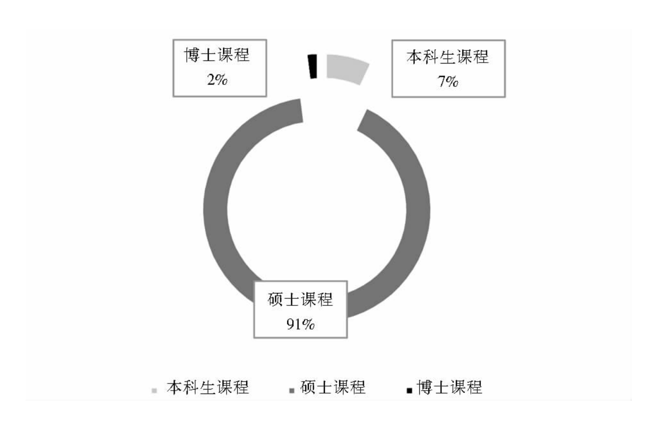


图 2 数据科学课程群培养对象类型

课程，分别是数据科学基础(Introduction to Data Science)、信息处理基础(Foundations of Information Processing)和数据库设计与原型构建(Database Design and Prototyping)。这些课程并没有要求学生提前拥有编程基础，而是通过课程为学生提供良好的数据库基础理论以及用编程语言解决抽象问题的方法，为学生在数据分析、数据科学、文本挖掘、数字图书馆和知识管理中的应用做好准备。UIUC 信息学院博士阶段同样没有特别详细的数据科学学位的培养目标和规划。现存的唯一一门面向博士生的数据科学课程，机器学习和社会计算前沿(Advanced Topics in Machine Learning & Social Computing)，是针对校内所有博士生进行教育的课程。课程内容主要涉及深度学习、生成对抗网络、对抗性学习、词嵌入以及经过筛选的人工智能中的热门主题(尤其是数据学习中的偏见、数据公平和数据伦理等)。博士阶段的数据科学课程采用研讨会的形式开展教学活动，学生们深入探讨以上主题的论文，在更广泛的理论、方法和领域中分析论文，并在自己的研究背景下对讨论的论文进行反思。

5.2 授课形式分析

UIUC 的授课形式有三种，分别是在校面授课程(On-Campus)、在线课程(Online)和混合式授课课程(On-Campus & Online)。UIUC 采用 Moodle 课程教学管理系统，帮助实现自主选课、教学信息发布、课程资料和课堂作业上传下载、群组互动研讨、学术报告宣讲等功能。2019 年以前，Moodle 系统通过借助 Blackboard Collaborate Ultra 网络会议系统可进行每周实时同步会话，提供双路音频视频、白板、分组讨论室及屏幕共享等服务；自 2020 年春季学期起，网络会议系统将逐步转换为 ZOOM 系统。

在 UIUC 信息学院数据科学课程群中，在校面授课程(On-Campus)占 55.1%；混合式授课课程(On-

Campus & Online）占 34.5%；在线课程（Online）占 10.3%，见表 2。由此可见，当前数据科学课程大多仍以传统的在校面授授课形式为主，如数据库管理及系统设计（Database Administration and Scaling for IS）和元数据理论与实践（Metadata in Theory and Practice）等。在科研第四范式——数据密集型科学范式的环境下，单纯地依靠传统在校面授的授课形式往往不能满足教学科研活动需求。在大数据时代，当数据科学课程群的培养对象为研究生群体时，新型的混合式授课模式在提升课堂质量、增强自身数据素养方面更能满足师生的需求，如数据挖掘（Data Mining）。混合式授课课程中部分课程采用实时同步的创新方式，方便学生自主选择学习时间、学习空间和学习方式，如数据清洗理论与实践（Theory and Practice Data Cleaning）。纯在线课程占比较小，大多集中在“数据科学 +”学科类课程中，如竞争情报与知识组织（Competitive Intelligence and Knowledge Management）。

表 2　UIUC iSchool 数据科学不同类型课程占比

	在校课程数（On-Campus）	在线课程数（Online）	在校 & 在线课程数（On-Campus & Online）	单类合计/占课程总比（%）
数据探索和准备	1	0	2	3/10.3
数据表示和转换	1	0	5	6/20.7
数据计算	1	1	0	2/6.9
数据建模	2	0	1	3/10.34
数据可视化和演示	4	0	0	4/13.8
数据科学 + 学科	7	2	2	11/37.9
总占比（%）	16/55.1	3/10.3	10/34.5	29/99.9

注：42 门课程中，仅有 29 门课程授课形式介绍完整

5.3　教师合作授课情况分析

独立授课指一门课程仅由一位教师进行教学活动（备课、授课、考核等）；合作授课指一门课程由两位或两位以上的教师和助教进行教学活动。按照独立授课和合作授课的定义为划分标准，笔者对 UIUC iSchool 的数据科学课程进行统计，得到如图 3 所示的数据科学课程群教师合作授课分析图。从整体来看，独立授课课程较多，共 16 门，合作授课课程较少，共 13 门，多数独立授课课程主要集中于“数据科学 + 学科”这一类，独占 8 门。除数据探索和准备类课程全部是合作授课，数据计算类课程全部是独立授课，其他类课程独立授课与合作授课相对均衡。

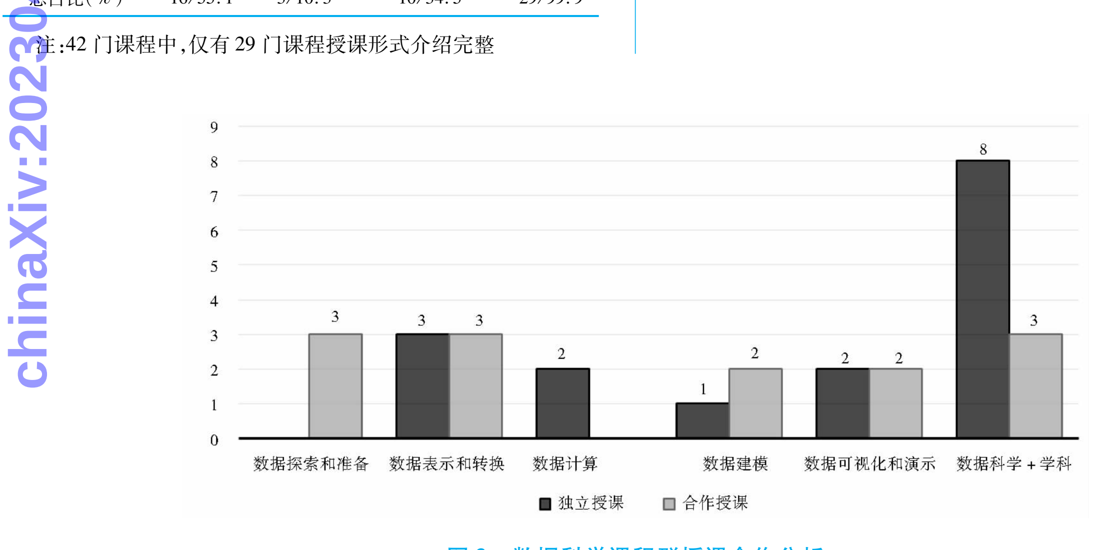


图 3　数据科学课程群授课合作分析

从图 3 可以发现六大类课程的授课合作情况各有特点。在不考虑部分课程内容信息缺失的情况下，数据探索和准备类课程中合作授课的课程有三门，这三门课程都采用在校面授的教学方式，且授课教师均为两人；其中数据挖掘课程（Data Mining）和信息处理基础课程（Foundations of Information Processing）采用在校面授 & 在线指导的混合式教学方式。数据表示和转换类课程中独立授课的课程和合作授课的课程均为三门，其中两门独立授课的课程，即开放数据融合课程（Open Data Mashups）和数据清洗理论与实践课程（Theory and Practice Data Cleaning）与一门合作授课的课程，即数据分析处理编程课程（Programming for Analytics and Data Processing）采用实时同步的新型教学方式，即教师在教室中的教学活动会实时更新到教学系统的指定栏目中，方便学生自主选择学习时间、学习空间和学习形式，为不同类型的学生接受课堂知识提供了更多的可能性。数据计算类课程全部为独立授课，且两门课程均采用在校面授的教学形式。云计算导论（Introduction to Cloud Computing）课程重点介绍云计算的各种服务应用场景，也讲解了公有云、私有云、混合

云、API 和数据安全等关键概念；机器学习和社会计算前沿(Advanced Topics in Machine Learning & Social Computing)课程主要通过学生积极参与深度学习、对抗性网络生成、对抗性学习以及人工智能某些话题的专题论文研讨会，深入分析这些专题论文的研究背景、研究理论、研究方法和研究成果。这两门课程均对教师迅速捕捉学生理解程度的能力有较高要求，故 UIUC 信息学院对数据计算类课程全部采用在校面授的教学形式，这种方式符合学生吸收知识、消化问题、反思感悟的现实规律，体现了该学院细致入微的学生教学设计和以"服务学生为本"的教学教育理念。

数据建模类课程中独立授课的课程有一门，合作授课的课程有两门，其中数据、统计模型和信息课程(Data, Statistical Models, and Information)由两名授课教师(Vetle Torvik 和 Jill Naiman)和三名助教(Chenyue Jiao, Xiaoliang Jiang 和 Pingjing Yang)组织完成教学，该课程是 UIUC 信息学院信息管理学硕士仅开设的三门必修课程之一，是进入本领域学习研究的基础。笔者在 UIUC 访学期间结合研究方向选择研修本门课程。该课程探讨的话题列举见表 3，包括数据模型信息概述，R 语言数据分析简介，可能性、条件关联性与贝叶斯定理，随机变量、期望和方差，数据推断基础，数值与分类数据，线性模型介绍，线性回归：多元线性回归，分类与逻辑回归。该课程首先回顾了概率论的相关内容，认真分析了常见的概率分布作为信息建模工具的优缺点；随后，介绍了包括参数和非参数预测模型，以及这些模型在无监督学习中的扩展。在所有的讨论中，本课程侧重于选择模型和度量模型的质量，及介绍统计概率模型在信息管理任务中的应用(例如，预测、排名和数据缩减等)。通过对课堂话题内容和课程简介内容的分析，可以看出数据、统计模型和信息课程(Data, Statistical Models, and Information)涉及统计学、机器学习、R 语言等理论与实践应用等多方面内容，教学任务繁重多样，教学信息体量巨大。若采用传统独立授课的教学形式，课程的教学质量难以得到保证，学生的积极性也无法得到激发。

表 3　数据、统计模型和信息课程课堂话题

课次	课堂话题	课堂话题翻译
1	Introductions and Overview of Data, Models, and Information	数据模型信息概述
2	Introduction to data analysis with R	R 语言数据分析简介
3	Probability; joint and conditional; Bayes Thm	可能性、条件关联性;贝叶斯定理
4	Random variables, expectation and variance	随机变量、期望和方差
5	Foundations for inference from data	数据推断基础
6	Numerical vs. Categorical data	数值与分类数据
7	Introduction to linear models; linear regression	线性模型介绍;线性回归
8	Linear regression: multiple linear regression	线性回归:多元线性回归
9	Classification and logistic regression	分类与逻辑回归

UIUC 信息学院信息管理学硕士学位必修课——数据、统计模型和信息课程(Data, Statistical Models, and Information)主要由副教授 V. Torvik 和客座教授 J. Naiman 联合采用合作授课的方式组织完成授课教学工作，两位授课教师的研究方向见表 4，其中 Vetle Torvik 主要从事于数学优化、计算统计、文本和数据挖掘、基于文献的发现和生物信息学领域的研究；客座教授 Jill Naiman 主要关注科学中有效且引人入胜的数据可视化的方法。两位授课教师分别在理论与实践两个方向上各有专长，合作授课可以更好地发挥各位授课教师的专长，更能提高该门信息管理学硕士学位必修课的课程品质。

表 4　数据、统计模型和信息课程授课教师研究方向及现阶段教授课程

姓名	职称	研究领域	现阶段教授课程
Vetle Torvik	Associate Professor 副教授	Mathematical optimization; computational statistics; text and data mining; literature-based discovery; bioinformatics.	Data Mining; Data, Statistical Models, and Information; Information Organization and Access; Methods for Data Science
Jill Naiman	Adjunct Lecturer 客座教授	methods for efficient and engaging data visualization in the sciences	Data Visualization; Data, Statistical Models, and Information; Foundations of Information Processing

UIUC 信息学院信息管理学硕士学位通过高度融合“人、信息和技术”的灵活课程，引领学生学会利用当下激增的信息资源来应对在组织和社会中面临的挑战，旨在塑造信息解决方案的专家。这说明了当下教学应注重合作授课，增加课程内容的丰富性和创新性才能满足对信息专业人才日益增长的市场需求。

数据可视化和演示类课程中独立授课的包括高级数据可视化（Advanced Data Visualization）和网络分析（Network Analysis）；合作授课的包括数据可视化（Data Visualization）和数据科学故事化（Data Science Storytelling）。基础类课程大多由多名教师合作完成授课，进阶类课程大多研究领域较为专深，故由该领域的专家独立授课效果更佳。数据科学 + 学科类中独立授课占比较高，因为该大类数据科学课程大多是针对某一具体问题的研究和探索进行介绍，实践性强，涉及学科种类广，与必修课、基础课相比学生受众范围较小。如数据伦理（Data Ethics）、实用健康数据分析（Practical Health Data Analytics）、商务分析（Business Analytics）和数字人文（Digital Humanities）等。

5.4 授课内容分析

笔者对数据科学课程群中所有课程简介的内容进行中文分词处理和关键词词频初步统计，删除全部数词、副词、介词、连词等无实际内容意义的词语，仅保留名词、动词和其他专有名词等，根据关键词权重指标（Score）得到排名前十五的关键词词汇，见表 5。关键词权重指标（Score）主要由关键词词频、IDF 倒转文档频率和关键词在文章中与其他词的语义聚合程度等决定[21]。通过分析课程内容的关键词词频，笔者发现现阶段 UIUC 数据科学课程较多涉及技术、可视化、概念介绍、建模与实践应用、数据分析、结构化、探讨学习、元数据、社交媒体、数据挖掘等重点内容。如数据可视化课程概述数据可视化的历史和运用的现代技术，这些将基于定量的、统计的和以网络为中心的数据集。课程主题包括交流可视化的建构、可视化的现代软件生态系统以及数据统计分析的可视化技术，尤其是关注 Python 生态系统和多维定量数据集。社交媒体分析课程主要向学生介绍社交媒体分析的基本概念、方法、技术和应用；培养学生分析结构化和非结构化社交媒体数据所需的素养和技能，以及有效、负责任地使用社交媒体分析的策略。通过循序渐进的指导，学生们将完成作业任务、动手练习和一项使他们能够分析来自各种现实世界平台（例如 Twitter 和维基百科）的用户生成数据的项目。

表 5 UIUC 数据科学课程群课程简介词频

序号	关键词	词频	权重	频率（%）
1	数据	68	1	0.139 630 4
2	技术	17	0.824 3	0.034 907 6
3	可视化	10	0.814 2	0.020 533 9
4	概念	12	0.796 2	0.024 640 7
5	科学	11	0.785 6	0.022 587 3
6	模型	9	0.775 6	0.018 480 5
7	工具	10	0.775 3	0.020 533 9
8	实践	9	0.764 8	0.018 480 5
9	理论	9	0.764 2	0.018 480 5
10	数据分析	7	0.747 7	0.014 373 7
11	结构化	6	0.746 2	0.012 320 3
12	数据科学	6	0.730 4	0.012 320 3
13	元数据	6	0.730 4	0.012 320 3
14	社交媒体	5	0.710 4	0.010 266 9
15	数据集	5	0.710 4	0.010 266 9

为进一步探析 UIUC 信息学院数据科学课程群中介绍讲解的编程语言、方法工具和技术应用领域，笔者从调研的 42 门课程简介中人工筛选出与工具、方法、编程语言、技能类等相关度较高的名词、外文字符和专有词汇为关键词，导入词云制作软件中进行直观的可视化演示，得到图 4。关键词词云，又称标签云，是以关键词字体的字号大小、颜色或是粗细来可视化各个关键词的重要程度，以便于读者快速把握文本信息的重要内容。通过对图 4 的分析，笔者发现 UIUC 信息学院数据科学课程群十分关注方法、工具、编程语言和软件等技能型知识的讲授与实践，主要包括 Python 语言、R 语言和 SQL（Structured Query Language，结构化查询语言）等。如数据科学基础（Foundations of Data Science）首先学习如何在 Unix 命令提示符下工作，随后介绍 Python 编程语言，重点介绍与数据科学相关的语言和相关 Python 模块的特定方面。Python 将主要通过 IPython 或 Jupyter 笔记本进行引入和使用，并将涵盖 Numpy、Scipy、MatPlotlib、Pandas、Seaborn 和 Scikit _ learn Python 模块。这些功能将通过简单的数据科学任务（如获取数据、清理数据、可视化数据和基本数据分析）进行演示。商务分析课程使用的工具主要包括 R、MySQL 和 Tableau。数据库类课程要求学生完成学习任务后精通结构化查询语言（SQL）编写基本查询语句，并全面了解关系数据库理论。将学习机器学习技术，包括监督和非监督学习、尺寸缩减和群集查找。重点将强调这些技术在高维数值数据、时间序列数据、图像数据和文本数据方面的实际应用。最后，学生将学

习使用关系数据库和云计算软件组件，如 Hadoop、Spark 和 NoSQL 数据存储。



图 4　数据科学课程群应用领域关键词

UIUC 信息学院数据科学课程除了重视方法、工具、编程语言等技能型知识的学习，还加强了对 git 和 GitHub 站点等源代码管理软件的学习。此外，这些技能型知识涉及的应用领域也比较广泛，包括隐私、通信、商业、法律、学术研究、图书馆、医疗卫生、政策标准道德、素养、社区、地理等。结合上文表 2，"数据科学＋"学科类课程占全部课程的比例最高也在此得到验证。

从国内外数据科学类岗位的招聘要求和相关著名数据科学家的访谈记录中可总结出，数据科学类岗位经常使用以下几种工具[22]：①R、Python、Haskell、Clojure、Scala 等数据科学语言工具；②NoSQL、MongoDB、Couchbase、Cassandra 等 NoSQL 工具；③ SQL、DW、RDMS、OLAP 等传统数据库和数据仓库工具；④HadoopHDFS&MapReduce、Spark 等支持大数据计算的工具；⑤Pig、HBase、Hive、Cascalog、Impala 等支持大数据管理、存储和查询的工具；⑥ Webscraper、Avro、Flume、Hume、Sqoop 等支持数据采集、聚合或传递的工具；⑦Pandas、SciPy、Weka、Knime 等支持数据挖掘的工具；⑧Tableu、Gephi、Shiny、D3. js、ggplot2 等数据可视化的工具；⑨SPSS、Matlab、SAS 等数据统计分析工具。

可见，UIUC 信息学院面向大数据时代的数据科学课程建设紧跟市场需求变化，涉及领域广泛，紧密关注数据的探索和准备、数据转换处理等技术、数据计算与建模工具的应用实践以及数据分析、数据可视化等数据技能建构，为全面培养满足社会需求的专业数据人才，优化数据科学教育提供良好的支撑。

6　总结与建议

UIUC 作为 iSchools 核心领导小组（iCaucus）的创始成员，一直致力于图情教育发展。面对科研第四范式带来的机遇和挑战，UIUC 已开始进行数据科学相关课程探索。本文通过实地学习和网站调研发现，UIUC 数据科学课程可分为六大类，面向本硕博各阶段学生，采用线上线下相结合的混合式教学方式，教学内容紧密跟随数据科学岗位市场要求。大多数类别课程均由教师合作授课。

目前国内数据科学教育体系与课程设置尚处于起步阶段，发展不成熟仍有许多需要提高的空间。本文结合上述对 UIUC 信息学院数据科学课程建设现状的调研，从培养对象类型、授课形式、授课合作程度和课程内容 4 个方面进行了归纳整理和对比分析，并针对当下我国数据科学教育面临的机遇和挑战提出以下 4 点建议：

（1）强化我国数据科学课程教育的连续性，使培养对象覆盖本硕博各阶段。经过广泛调研发现，现阶段我国大多数高校的数据科学课程教育集中在本科阶段，硕士研究生阶段的培养较为初步，博士研究生的培养更是少之又少。如 5.1 所述，UIUC 现阶段数据科学硕士课程建设和培养体系较为成熟，博士阶段课程采用研讨会的形式开展，本科阶段课程仅仅旨在为部分核心专业课提供良好的数据库理论基础，仍在不断建设完善中。因此，未来国内高校需以提供覆盖本硕博各阶段的数据科学系统教育和课程建设为发展目标。

（2）丰富我国数据科学课程教学的创新性，使混合式教学法进一步融入国内课程教学实践。如 5.2 所述，UIUC 信息学院主要采用多功能的 Moodle 教学管理系统，据了解 2020 年 UIUC 将升级为 ZOOM，届时功能将更加强大。现阶段国内 MOOC 等平台上不乏优质精品在线开放课程，与数据科学相关的课程包含中国人民大学朝乐门教授的数据科学导论、北京理工大学嵩天教授的 Python 数据分析与展示等。利用 MOOC 改进教学，尝试开设精品在线课程，学习 UIUC 信息学院将课程教学管理平台融入日常教学活动，仍是我国高校未来需要不断改进的方向。

（3）增强我国数据科学授课教师间的教学合作，使授课合作程度进一步深化。如 5.3 所述，UIUC 信息学院数据科学课程群中数据探索和准备类课程全部是合作授课，并且该校数据科学课程群授课教师除了来自传统图书情报专业，部分教师还拥有计算机科学专业背景和图书馆数据管理业务实践背景。近年国内各大高校也逐渐将图书馆员请进课堂，共同进行教学设计完成教学任务；但院系内部任课教师间的教学合作

仍有待加强,师资引进方面也需增强跨学科视角,根据实际业务需要聘用复合型人才,可增加兼职教授和客座教授数量。

(4)完善我国数据科学课程研究方向,使课程内容紧密跟随市场需求。国内数据科学课程大多以计算机科学和统计学为背景,而以图书馆数据实践业务、医学信息分析业务等为主导方向的课程很少。数据科学课程对于数据素养能力提升和数据思维培养,以及大数据时代新兴高技能知识人才——数据馆员专业性的保障都具有非同寻常的意义[23]。如 5.4 所述,UIUC 信息学院数据科学课程涉及领域广泛,紧密关注信息行业图情领域应用实践和数据技能建构,为培养专业数据人才提供了良好的支撑。因此,我国高校应深入调研国内信息市场岗位要求,在参考国外一流院校的优秀课程内容和设计基础上,将各种实用工具及语言有选择地融入课程实验并组织学生上机操作。

参考文献:

[1] 中华人民共和国工业和信息化部. 大数据产业发展规划(2016-2020 年)[EB/OL]. [2019-10-21]. http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5464999/content.html.

[2] Talkingdata. 专业数据人才教育行业生态报告[R/OL]. [2019-08-30]. http://mi.talkingdata.com/report-detail.html? id=750.

[3] 中华人民共和国教育部. 深入推进"新工科"建设[EB/OL]. [2019-11-27]. http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2606/2019/tqh20191031/sfcl/201910/t20191031_406260.html.

[4] NAUR P. Concise survey of computer methods[M]. Lund, Sweden:Student Litteratur, 1974:1-30.

[5] CONWAY D. The data science venn diagram[EB/OL]. [2019-08-31]. http://drewconway.com/zia/ 2013/3/26/the-data-science-venn-diagram.

[6] ANDERSON P, BOWRING J. An undergraduate degree in data science: curriculum and a decade of implementation experience [EB/OL]. [2019-10-21]. https://blogs.valpo.edu/datadesk/files/ 2017/04/Charleston-SIGCSE14-designofDataScience-withObjectives.pdf.

[7] BAUMER B, COLLEGE S. A data science course for undergraduates: thinking with Data[J]. The American statistician, 2015, 69(4):334-342.

[8] VEAUX R, AGARWAL M. Curriculum guidelines for undergraduate programs in data science [EB/OL]. [2019-10-21]. https://www.stat.berkeley.edu/~nolan/Papers/Data.Science.Guidelines.16.9.25.pdf.

[9] SONG V, ZHU Y. Big data and data science: opportunities and challenges of iSchools[J]. Journal of data and information science, 2017,2(3):1-18.

[10] SONG V, ZHU Y. Big data and data science: what should we teach? [J]. Expert systems, 2016,33(4):364-373.

[11] 朝乐门,杨灿军,王盛杰,等. 全球数据科学课程建设现状的实证分析[J]. 数据分析与知识发现,2017,1(6):12-21.

[12] 朝乐门,邢春晓,王雨晴. 数据科学与大数据技术专业特色课程研究[J]. 计算机科学,2018,3(45):3-10.

[13] 李莎莎,周竞文,唐晋韬,等. 数据科学与大数据人才专业课程体系分析[J]. 计算机工程与科学,2018,(40):109-113.

[14] 闫慧,张钰浩,张鑫灿,等. iSchools 联盟数据科学教育项目现状调查[J]. 情报资料工作,2018(4):95-100.

[15] 苏日娜,杨沁. LIS 学科中数据科学课程体系设置研究——以 iSchools 高校课程调研为中心[J]. 图书馆论坛,2019,39(4):40-49.

[16] 邓胜利,付少雄. 计算科学与信息科学融合下的图书情报学科建设[J]. 情报资料工作,2019,40(3):80-87.

[17] SONG V, ZHU Y. Big data and data science: opportunities and Challenges of iSchools[J]. Journal of data and information science, 2017,2(3):1-18.

[18] DHAR V. Data science and prediction[J]. Communications of the ACM, 2013, 56(12): 64-73.

[19] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making[J]. Big data, 2013, 1(1):51-59.

[20] DONOHO D. 50 years of data science——a presentation at the Tukey Centennial workshop[R/OL]. [2020-05-30]. http://www.mathsccnu.com/forum.php? mod=attachment&aid=MzQwN3xkMGE3YWQ3NnwxNDQ1NTI3Mjg1fDF8MzkwMA%3D%3D.

[21] 官方权威发布"图悦"热词分析指标说明[EB/OL]. [2019-11-24]. https://mp.weixin.qq.com/s? __biz=MjM5MDAzOTk4OA==&mid=204190927&idx=1&sn=079a1b63a5a8d1a926307957d954d758&scene=1&from=groupmessage&isappinstalled=0#rd.

[22] 朝乐门. 数据科学[M]. 北京:清华大学出版社,2016:31-32.

[23] 魏来,高希然. 大数据背景下高校数据馆员的角色定位[J]. 情报资料工作,2015(5):90-94.

作者贡献说明:

杨瑞仙:构思、收集资料、审核与修改论文;

万佳琦:收集、整理分析研究资料,撰写与修改论文。

Investigation Research on Construction of Data Science Courses in UIUC iSchool

Yang Ruixian Wan Jiaqi
School of Information Management, Zhengzhou University, Zhengzhou 450001

Abstract: [Purpose/significance] This paper studied the current construction of data science curriculum groups, focuses on the training program of data science talents and provides references and advice for the data science practice of information colleges in China. [Method/process] Based on the data science curriculum practice of UIUC iSchool, this paper first investigated the name and introductions of the data science-related courses in detail, academic hours, teaching forms, the teachers and the subjects, then systematically classified the course groups and made a detailed comparative analysis from the 4 aspects: training object type, teaching forms, teaching cooperation degree, and the course content. Finally, this paper summarized the enlightenments and suggestions for the development of data science education in China in light of the current domestic situation. [Result/conclusion] In UIUC iSchool Data science courses can be divided into 6 categories, which are suitable for students at all stages. A mixed teaching method combining online and offline is adopted. Teachers cooperate with each other and the contents closely follow the data science job market requirements. Finally, the authors suggest that we should strengthen the continuity of cultivation, innovate teaching methods, improve teaching cooperation, and enrich research directions in the field of data science in China.

Keywords: data science UIUC iSchool curriculum construction curriculum reform